

Informatique théorique et Applications/Theoretical Informatics and Applications  
(vol. 24, n° 6, 1990, p. 521 à 530)

## ON DOT-DEPTH TWO (\*)

by F. BLANCHET-SADRI

Communicated by A. ARNOLD

**Abstract.** — For positive integers  $m_1, \dots, m_k$ , congruences  $\sim_{(m_1, \dots, m_k)}$  related to a version of the Ehrenfeucht-Fraïssé game are defined which correspond to level  $k$  of the Straubing hierarchy of star-free languages. Given any finite alphabet  $A$ , a necessary and sufficient condition is given for the monoids  $A^*/\sim_{(m_1, m_2, m_3)}$  to be of dot-depth exactly 2.

**Résumé.** — Étant donné des entiers positifs  $m_1, \dots, m_k$ , on définit des congruences  $\sim_{(m_1, \dots, m_k)}$  en relation avec une version du jeu de Ehrenfeucht-Fraïssé, et qui correspondent au niveau  $k$  de la hiérarchie de concaténation de Straubing. Étant donné un alphabet fini  $A$ , une condition nécessaire et suffisante est donnée pour que les monoides définis par ces congruences soient de dot-depth exactement 2.

### 1. INTRODUCTION

Let  $A$  be a given finite alphabet. The regular languages over  $A$  are those subsets of  $A^*$ , the free monoid generated by  $A$ , constructed from the finite languages over  $A$  by the boolean operations, the concatenation product and the star. The star-free languages are those regular languages which can be obtained from the finite languages by the boolean operations and the concatenation product only. According to Schützenberger [15],  $L \subseteq A^*$  is star-free if and only if its syntactic monoid  $M(L)$  is finite and aperiodic. General references on the star-free languages are McNaughton and Papert [10], Eilenberg [6] or Pin [12].

Natural classifications of the star-free languages are obtained based on the alternative use of the boolean operations and the concatenation product. Let  $A^+ = A^* \setminus \{1\}$ , where 1 denotes the empty word. Let

(\*) Received in January 1989, revised in February 1989.

(<sup>1</sup>) Department of Mathematics, McGill University, 805 Sherbrooke Street West, Montreal, P. Quebec, Canada, H3A 2K6.

$$A^+ \mathcal{B}_0 = \{ L \subseteq A^+ \mid L \text{ is finite or cofinite} \},$$

$$A^+ \mathcal{B}_{k+1} = \{ L \subseteq A^+ \mid L \text{ is a boolean combination of languages of the form } L_1 \dots L_n (n \geq 1) \text{ with } L_1, \dots, L_n \in A^+ \mathcal{B}_k \}.$$

Only nonempty words over  $A$  are considered to define this hierarchy; in particular, the complement operation is applied with respect to  $A^+$ . The language classes  $A^+ \mathcal{B}_0, A^+ \mathcal{B}_1, \dots$  form the so-called dot-depth hierarchy introduced by Cohen and Brzozowski in [4]. The union of the classes  $A^+ \mathcal{B}_0, A^+ \mathcal{B}_1, \dots$  is the class of star-free languages.

Our attention is directed toward a closely related and more fundamental hierarchy, this one in  $A^*$ , introduced by Straubing in [18]. Let

$$A^* \mathcal{V}_0 = \{ 0, A^* \},$$

$$A^* \mathcal{V}_{k+1} = \{ L \subseteq A^* \mid L \text{ is a boolean combination of languages of the form } L_0 a_1 L_1 a_2 \dots a_n L_n (n \geq 0) \text{ with } L_0, \dots, L_n \in A^* \mathcal{V}_k \text{ and } a_1, \dots, a_n \in A \}.$$

$L \subseteq A^*$  is star-free if and only if  $L \in A^* \mathcal{V}_k$  for some  $k \geq 0$ . The *dot-depth* of  $L$  is the smallest such  $k$ .

Using Eilenberg's correspondence, we have that for each  $k \geq 0$ , there is a variety  $V_k$  of finite monoids such that for  $L \subseteq A^*$ ,  $L \in A^* \mathcal{V}_k$  if and only if  $M(L) \in V_k$ . An outstanding open problem is whether one can decide if a language has dot-depth  $k$ , i.e., can we effectively characterize the varieties  $V_k$ ? The variety  $V_0$  consists of the trivial monoid alone,  $V_1$  of all finite  $\mathcal{T}$ -trivial monoids [16]. Straubing [19] conjectured an effective characterization, based on the syntactic monoid of the language, for the case  $k=2$ . His characterization, formulated in terms of a novel use of categories in semigroup theory recently developed by Tilson [22], is shown to be necessary in general, and sufficient for an alphabet of two elements.

In the framework of semigroup theory, Brzozowski and Knast [1] showed that the dot-depth hierarchy is infinite. Thomas [21] gave a new proof of this result, which shows also that the Straubing hierarchy is infinite, based on a logical characterization of the dot-depth hierarchy that he obtained in [20] (Perrin and Pin gave one for the Straubing hierarchy [11]) and the following version of the Ehrenfeucht-Fraïssé game.

First, one regards a word  $w \in A^*$  of length  $|w|$  as a word model  $w = \langle \{1, \dots, |w|\}, <^w, (Q_a^w)_{a \in A} \rangle$  where the universe  $\{1, \dots, |w|\}$  represents the set of positions of letters in  $w$ ,  $<^w$  denotes the  $<$ -relation in  $w$ , and  $Q_a^w$  are unary relations over  $\{1, \dots, |w|\}$  containing the positions with letter  $a$ , for each  $a \in A$ . For a sequence  $\bar{m} = (m_1, \dots, m_k)$  of positive integers, where  $k \geq 0$ , the game  $\mathcal{G}_{\bar{m}}(u, v)$  is played between two players I and II on the word models  $u$  and  $v$ . A play of the game consists of  $k$  moves. In the

$i$ -th move, player I chooses, in  $u$  or in  $v$ , a sequence of  $m_i$  positions; then player II chooses, in the remaining word, also a sequence of  $m_i$  positions. After  $k$  moves, by concatenating the sequences chosen from  $u$  and  $v$ , two sequences  $p_1 \dots p_n$  from  $u$  and  $q_1 \dots q_n$  from  $v$  have been formed where  $n = m_1 + \dots + m_k$ .

Player II has won the play if

$$p_i <^u p_j \quad \text{if and only if} \quad q_i <^v q_j, \quad (1)$$

and

$$Q_a^u p_i \quad \text{if and only if} \quad Q_a^v q_i, \quad a \in A \quad \text{for } 1 \leq i, j \leq n. \quad (2)$$

If there is a winning strategy for II in the game  $\mathcal{G}_{\bar{m}}(u, v)$  to win each play we write  $u \sim_{\bar{m}} v$ .  $\sim_{\bar{m}}$  naturally defines a congruence on  $A^*$  which we denote also by  $\sim_{\bar{m}}$ . The standard Ehrenfeucht-Fraïssé game [5] is the special case  $\mathcal{G}_{(1, \dots, 1)}(u, v)$ . Thomas [20], [21] and Perrin and Pin [11] imply that  $L \in A^* \mathcal{V}_k$  if and only if  $L$  is a  $\sim_{\bar{m}}$ -language for some  $\bar{m} = (m_1, \dots, m_k)$  (or  $L$  is a union of classes of the congruence  $\sim_{\bar{m}}$ ). This congruence characterization implies that the problem of deciding whether a language has dot-depth  $k$  is equivalent to the problem of effectively characterizing the monoids  $M = A^* / \sim$  with  $\sim \supseteq \sim_{\bar{m}}$  for some  $\bar{m} = (m_1, \dots, m_k)$ , i.e.,

$$V_k = \{ A^* / \sim \mid \sim \supseteq \sim_{\bar{m}} \text{ for some } \bar{m} = (m_1, \dots, m_k) \}.$$

This paper is concerned with an application of the above congruence characterization. We show that  $A^* / \sim_{(m_1, m_2, m_3)}$  is of dot-depth exactly 2 if and only if  $m_2 = 1$ . The proof relies on some properties of the congruences  $\sim_{\bar{m}}$  stated in the next section. [2] and [3] include other applications: among them are an answer to a conjecture of Pin [13] concerning tree hierarchies of monoids and also systems of equations satisfied in natural sublevels of level 1 of the Straubing hierarchy. The reader is referred to the books by Eilenberg [6], Lallement [9], Pin [12], Enderton [7] and Fraïssé [8] for all the algebraic and logical terms not defined here.

## 2. SOME PROPERTIES OF THE CHARACTERIZING CONGRUENCES

### 2.1. An induction lemma

The following lemma is a basic result (similar to one in [14] regarding  $\sim_{(1, \dots, 1)}$ ) which allows to resolve games with  $k+1$  moves into games with

$k$  moves and thereby allows to perform induction arguments. In what follows,  $u^{<p}(u_{>p})$  denotes the subword of  $u$  to the left (right) of position  $p$  and  $u_{>p}^{<q}$  the subword of  $u$  between positions  $p$  and  $q$ .

LEMMA 2.1.: Let  $\bar{m} = (m_1, \dots, m_k)$ .  $u \sim_{(\bar{m}, m_1, \dots, m_k)} v$  if and only if

(1) for every  $p_1, \dots, p_m \in u$  ( $p_1 \leq \dots \leq p_m$ ) there are  $q_1, \dots, q_m \in v$  ( $q_1 \leq \dots \leq q_m$ ) such that

(i)  $Q_a^u p_i$  if and only if  $Q_a^v q_i$ ,  $a \in A$  for  $1 \leq i \leq m$ ,

(ii)  $u^{<p_1} \sim_{\bar{m}} v^{<q_1}$ ,

(iii)  $u_{>p_i+1}^{<p_{i+1}} \sim_{\bar{m}} v_{>q_i+1}^{<q_{i+1}}$  for  $1 \leq i \leq m-1$ ,

(iv)  $u_{>p_m} \sim_{\bar{m}} v_{>q_m}$  and

(2) for every  $q_1, \dots, q_m \in v$  ( $q_1 \leq \dots \leq q_m$ ) there are  $p_1, \dots, p_m \in u$  ( $p_1 \leq \dots \leq p_m$ ) such that (i), (ii), (iii) and (iv) hold.

## 2.2. A lemma for inclusion

Define

$$\mathcal{N}_{(m_1, \dots, m_k)} = m_1 + \dots + m_k + \sum_{1 \leq i_1 < i_2 \leq k} m_{i_1} m_{i_2} + \dots + \sum_{1 \leq i_1 < \dots < i_{k-1} \leq k} m_{i_1} \dots m_{i_{k-1}} + m_1 \dots m_k.$$

One can show that  $x^N \sim_{(\bar{m}, \dots, m_k)} x^{N+1}$  ( $N = \mathcal{N}_{(m_1, \dots, m_k)}$ ) and that  $N$  is the smallest  $n$  such that  $x^n \sim_{(\bar{m}, \dots, m_k)} x^{n+1}$  (the proof is similar to the one of a property of  $\sim_{(1, \dots, 1)}$  in [21]). We see that if  $u, v \in A^*$  and  $u \sim_{(\bar{m}, \dots, m_k)} v$ , then  $|u|_a = |v|_a < \mathcal{N}_{(m_1, \dots, m_k)}$  or  $|u|_a, |v|_a \geq \mathcal{N}_{(m_1, \dots, m_k)}$  (here  $|w|_a$  denotes the number of occurrences of the letter  $a$  in  $w$ ). The following lemma follows easily from Lemma 2.1 and the above remarks.

LEMMA 2.2 :

$$\sim_{(m_1, \dots, m_k)} \subseteq \sim_{(\mathcal{N}_{(m_1, \dots, m_k)}, \dots, m_k)} \quad \text{and} \quad \sim_{(m_1, \dots, m_k)} \not\subseteq \sim_{(\mathcal{N}_{(m_1, \dots, m_k)}+1, \dots, m_k)}.$$

If  $k \leq k'$  and  $\exists 0 = j_0 < \dots < j_{k-1} < j_k = k'$  such that  $m_i \leq \mathcal{N}_{(m'_{j_{i-1}+1}, \dots, m'_{j_i})}$  for  $1 \leq i \leq k$ , then  $\sim_{(m'_1, \dots, m'_{k'})} \subseteq \sim_{(m_1, \dots, m_k)}$ .

## 3. A SEQUENCE OF MONOIDS OF DOT-DEPTH 2

In this section, we show that for positive integers  $m_1, m_2$  and  $m_3$ ,  $A^*/\sim_{(m_1, m_2, m_3)}$  is of dot-depth exactly 2 if and only if  $m_2 = 1$ . The following lemma shows the necessity of the condition.

LEMMA 3.1: Let  $m_1$  and  $m_3$  be positive integers. Then  $A^*/\sim_{(m_1, 2, m_3)}$  is of dot-depth exactly 3.

*Proof:* Let  $m > 0$ . Consider  $u_m = ((xy)^m x (xy)^{2m} y (xy)^m)^m$ ,  $v_m = ((xy)^m y (xy)^{2m} x (xy)^m)^m$ . A result of Straubing [17] implies that monoids in  $V_2$  are 2-mutative and hence satisfy  $u_m = v_m$  for all sufficiently large  $m$ . However, for every  $N \geq \mathcal{N}_{(1, 2, 1)}$ ,  $u_N x_{(1, 2, 1)} v_N$ . To see this, we illustrate a winning strategy for player I in the game  $\mathcal{G}_{(1, 2, 1)}(u_N, v_N)$ . (I,  $i$ ) denotes a position chosen by player I in the  $i$ -th move,  $i = 1, 2, 3$ . Similarly, (II,  $i$ ) denotes a position chosen by player II in the  $i$ -th move. Player I, in the first move, chooses the

$$\begin{array}{c} \text{2N} \qquad \qquad \qquad \text{N} \\ \text{u}_N = \dots (xy)^N x \overbrace{(xy)(xy) \dots (xy)(xy)}^{2N} \overbrace{y(xy)(xy) \dots (xy)(xy)}^N \\ \qquad \qquad \qquad \uparrow \qquad \qquad \qquad \uparrow \quad \uparrow \\ \qquad \qquad \qquad \text{(II, 1)} \qquad \qquad \qquad \text{(I, 2)} \end{array}$$

$$\begin{array}{c} \text{2N} \qquad \qquad \qquad \text{N} \\ \text{v}_N = \dots (xy)^N x \overbrace{(xy)(xy) \dots (xy)(xy)}^{2N} \overbrace{y(xy)(xy) \dots (xy)(xy)}^N \\ \qquad \qquad \qquad \uparrow \qquad \qquad \qquad \uparrow \quad \uparrow \\ \qquad \qquad \qquad \text{(I, 1)} \qquad \qquad \qquad \text{(II, 2)} \end{array}$$

last  $x$  followed immediately by an  $x$  in  $v_N$ . Player II, in the first move, has to choose the last  $x$  followed immediately by an  $x$  in  $u_N$  (if not, player I in the next two moves could win by choosing in the second move the last two consecutive  $x$ 's in  $u_N$ ). Player I, in the second move, chooses the last two consecutive  $y$ 's in  $u_N$ . Player II, in the second move, cannot choose two consecutive  $y$ 's in  $v_N$  to the right of the previously chosen position. Hence he is forced to choose two  $y$ 's separated by an  $x$ . Player I, in the third move, selects that  $x$ . But player II loses since he cannot choose an  $x$  between the two consecutive  $y$ 's chosen in the preceding move by I. The result follows. [ ]

Assume  $|u|_a, |v|_a > 0$ . Let  $u = u_0 a u_1 \dots a u_{|u|_a}$ ,  $v = v_0 a v_1 \dots a v_{|v|_a}$ . If  $Q_a^u p_i, Q_a^v q_j$  for  $i = 1, \dots, |u|_a$ ,  $j = 1, \dots, |v|_a$ , then  $u_i = u_{>p_i+1}^{<p_{i+1}}$ ,  $i = 1, \dots, |u|_a - 1$ ,  $v_j = v_{>q_j+1}^{<q_{j+1}}$ ,  $j = 1, \dots, |v|_a - 1$ .  $u_0 = u^{<p_1}$ ,  $v_0 = v^{<q_1}$ ,  $u_{|u|_a} = u_{>p_{|u|_a}}$ ,  $v_{|v|_a} = v_{>q_{|v|_a}}$ .

The next two lemmas will be used in showing that for positive integers  $m_1$  and  $m_3$ ,  $A^*/\sim_{(m_1, 1, m_3)}$  is of dot-depth exactly 2.



a position in  $u^{<p}$  (the proof is similar when starting with a position in  $v^{<q}$ ). Assume  $Q_{a_i}^u p'$ .

Case 1:  $p'$  is among the first  $m'_2$  occurrences of  $a_i$  in  $u^{<p}$ .

Let  $q'$  be the same occurrence among the first  $m'_2$  occurrences of  $a_i$  in  $v^{<q}$ . It is clear that  $u^{<p}_{>p'} \sim_{(m'_2)} v^{<q}_{>q'}$  and  $u^{<p'} \sim_{(m'_2)} v^{<q'}$ .

Case 2:  $p'$  is among the last  $m'_2$  occurrences of  $a_i$  in  $u^{<p}$ . Similar to case 1.

Case 3:  $p'$  is not among the first  $m'_2$  nor the last  $m'_2$  occurrences of  $a_i$  in  $u^{<p}$ .

Let  $p''$  and  $p'''$  ( $p'' < p'''$ ) be the closest positions to  $p'$  in  $u^{<p'}$  and  $u^{<p}_{>p'}$ , respectively among the chosen positions by player I. Let  $q''$  and  $q'''$  ( $q'' < q'''$ ) be the corresponding positions chosen by player II.

Since  $u^{<p'''} \sim_{(m'_2)} v^{<q'''}_{>q'''}$ , there is  $q'$  in  $v^{<q'''}_{>q'''}$  such that  $Q_{a_i}^v q'$ .

Let us show that  $u^{<p}_{>p'} \sim_{(m'_2)} v^{<q}_{>q'}$ .  $u^{<p'} \sim_{(m'_2)} v^{<q'}$  follows similarly.

Let  $w = w_1 \dots w_{|w|}$ ,  $|w| \leq m'_2$  in  $v^{<q}_{>q'}$ . The proof is similar when starting with  $w$  in  $u^{<p}_{>p'}$ . If  $w \in v^{<q}_{>q'}$ , it is clear that  $w \in u^{<p}_{>p'}$ , hence in  $u^{<p}_{>p'}$ . So let us assume  $w \notin v^{<q}_{>q'}$ . Let  $p_{w_1}, \dots, p_{w_{|w|}}$  in  $v^{<q}_{>q'}$ , at least  $p_{w_1}$  being in  $v^{<q}_{>q'}$ , which spell  $w_1 \dots w_{|w|}$ .  $p_{w_1}, \dots, p_{w_{|w|}}$  are hence positions which spell an occurrence of a subword of length  $\leq m'_2$  in  $v^{<q}_{>q'}$ . Hence they are smaller than or equal to those positions which spell the last occurrence of  $w$  in  $v^{<q}_{>q'}$  which are in  $v^{<q}_{>q'}$ . Hence  $w \in u^{<p}_{>p'}$ . [ ]

The following corollary gives another result for inclusion (one was Lemma 2.2).

COROLLARY 3.5: Let  $|A| = r$ . Then

$$\sim_{(m'_1 + (m'_1 + 1) 2m'_2 (r+1)^{m'_2}, m'_2)} \subseteq \sim_{(m'_1, \mathcal{N}_{(1, m'_2)})}$$

Proof: From Theorem 3.4 and Lemma 2.2. [ ]

#### ACKNOWLEDGMENTS

This research was partially supported by F.C.A.R. (Quebec). We would like to thank the referees for their comments and suggestions.

#### REFERENCES

1. J. A. BRZOWSKI and R. KNAST, The Dot-Depth Hierarchy of Star-Free Languages if Infinite, *J. Comp. Sys. Sci.*, 1978, 16, pp. 37-55.

2. F. BLANCHET-SADRI, Some Logical Characterizations of the Dot-Depth Hierarchy and Applications, *Technical Report No.88-03 of the Department of Mathematics and Statistics of McGill University*, July 1988, pp. 1-44.
3. F. BLANCHET-SADRI, Games, Equations and the Dot-Depth Hierarchy, (preprint 1988), *Computers and Mathematics with applications* (à paraître).
4. R. S. COHEN and J. A. BRZOWSKI, Dot-Depth of Star-Free Events, *J. Comp. Sys. Sci.*, 1971, 5, pp. 1-16.
5. A. EHRENFUCHT, An Application of Games to the Completeness Problem for Formalized Theories, *Fund. Math.*, 1961, 49, pp. 129-141.
6. S. EILENBERG, Automata, Languages and Machines, B, *Academic Press*, New York, 1976.
7. H. B. ENDERTON, A Mathematical Introduction to Logic, *Academic Press*, New York, 1972.
8. R. FRAISSÉ, Cours de logique mathématique, tome 2, Gauthier-Villars, Paris, 1972.
9. G. LALLEMENT, Semigroups and Combinatorial Applications, *Wiley*, New York, 1979.
10. R. MCNAUGHTON and S. PAPERT, Counter-Free Automata, *M.I.T. Press*, Cambridge, Mass., 1971.
11. D. PERRIN and J. E. PIN, First-Order Logic and Star-Free Sets, *J. Comp. Sys. Sci.*, 1986, 32, pp. 393-406.
12. J. E. PIN, Variétés de langages formels, *Masson*, Paris, 1984.
13. J. E. PIN, Hiérarchies de contaténation, *R.A.I.R.O. Informatique Théorique*, 1984, 18, pp. 23-46.
14. J. G. ROSENSTEIN, Linear Orderings, *Academic Press*, New York, 1982.
15. M. P. SCHÜTZENBERGER, On Finite Monoids having only Trivial Subgroups, *Information and Control*, 1965, 8, pp. 190-194.
16. I. SIMON, Piecewise Testable Events, Proc. 2nd GI Conference, *Lectures Notes in Comput. Sci.*, Springer Verlag, Berlin, 1975, 33, pp. 214-222.
17. H. STRAUBING, A Generalization of the Schützenberger Product of Finite Monoids, *Theoretical Comput. Sci.*, 1981, 13, pp. 137-150.
18. H. STRAUBING, Finite Semigroup Varieties of the Form  $V * D$ , *J. of Pure and Applied Algebra*, 1985, 36, pp. 53-94.
19. H. STRAUBING, Semigroups and Languages of Dot-Depth Two, *Proc. 13th ICALP, Lecture Notes in Comput. Sci.*, Springer Verlag, New York, 1986, 226, pp. 416-423.
20. W. THOMAS, Classifying Regular Events in Symbolic Logic, *J. Comp. Sys. Sci.*, 1982, 25, pp. 360-376.
21. W. THOMAS, An Application of the Ehrenfeucht-Fraissé Game in Formal Language Theory, *Bull. Soc. Math. de France*, 2<sup>e</sup> série, Mémoire, 1984, No. 16, pp. 11-21.
22. B. TILSON, Categories as Algebra, *J. of Pure and Applied Algebra*, 1987, 48, pp. 83-198.